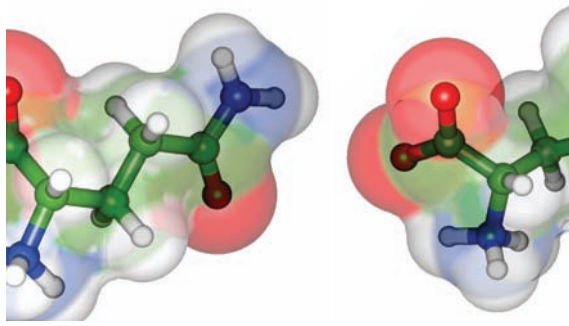


Amino Acids

CHAPTER 4

1 **The Amino Acids of Proteins**
- A. General Properties
- B. Peptide Bonds
- C. Classification and Characteristics
- D. Acid–Base Properties
- E. A Few Words on Nomenclature

2 **Optical Activity**
- A. An Operational Classification
- B. The Fischer Convention
- C. The Cahn–Ingold–Prelog System
- D. Chirality and Biochemistry

3 **"Nonstandard" Amino Acids**
- A. Amino Acid Derivatives in Proteins
- B. Specialized Roles of Amino Acids

It is hardly surprising that much of the early biochemical research was concerned with the study of proteins. Proteins form the class of biological macromolecules that have the most well-defined physicochemical properties, and consequently they were generally easier to isolate and characterize than nucleic acids, polysaccharides, or lipids. Furthermore, proteins, particularly in the form of enzymes, have obvious biochemical functions. The central role that proteins play in biological processes has therefore been recognized since the earliest days of biochemistry. In contrast, the task of nucleic acids in the transmission and expression of genetic information was not realized until the late 1940s and their catalytic function only began to come to light in the 1980s, the role of lipids in biological membranes was not appreciated until the 1960s, and the biological functions of polysaccharides are still somewhat mysterious.

In this chapter we study the structures and properties of the monomeric units of proteins, the **amino acids.** It is from these substances that proteins are synthesized through processes that we discuss in Chapter 32. Amino acids are also energy metabolites and, in animals, many of them are essential nutrients (Chapter 26). In addition, as we shall see, many amino acids and their derivatives are of biochemical importance in their own right (Section 4-3B).

1 THE AMINO ACIDS OF PROTEINS

The analyses of a vast number of proteins from almost every conceivable source have shown that *all proteins are composed of the 20 "standard" amino acids listed in Table 4-1.* These substances are known as **α-amino** acids because, with the exception of **proline,** they have a primary amino group and a carboxylic acid group substituent on the same carbon atom (Fig. 4-1; proline has a secondary amino group).

A. *General Properties*

The pK values of the 20 "standard" α-amino acids of proteins are tabulated in Table 4-1. Here pK_1 and pK_2, respectively, refer to the α-carboxylic acid and α-amino groups, and pK_R refers to the side groups with acid–base properties. Table 4-1 indicates that the pK values of the α-carboxylic acid groups lie in a small range around 2.2 so that above pH 3.5 these groups are almost entirely in their carboxylate forms. The α-amino groups all have pK values near 9.4 and are therefore almost entirely in their ammonium ion forms below pH 8.0. This leads to an important structural point: *In the physiological pH range, both the carboxylic acid and the amino groups of α-amino acids are completely ionized* (Fig. 4-2). An amino acid can therefore act as either an acid or a base. Substances with this property are said to be **amphoteric** and are referred to as **ampholytes** (*ampho*teric electro*lytes*). In Section 4-1D, we shall delve a bit deeper into the acid–base properties of the amino acids.

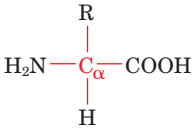


Figure 4-1 **General structural formula for α-amino acids.** There are 20 different R groups in the commonly occurring amino acids (Table 4-1).

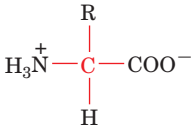
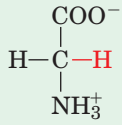
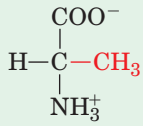
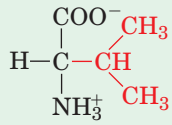
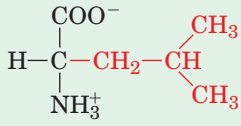
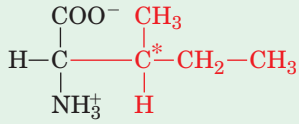
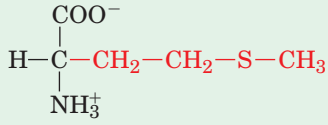
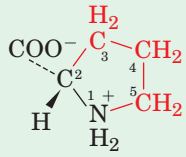
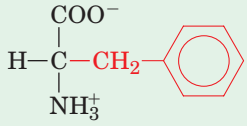
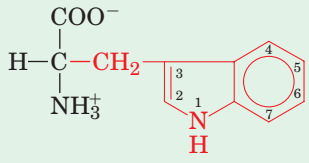


Figure 4-2 **Zwitterionic form of the α-amino acids that occurs at physiological pH values.**

Table 4-1 **Covalent Structures and Abbreviations of the "Standard" Amino Acids of Proteins, Their Occurrence, and the p*K* Values of Their Ionizable Groups**

Name, Three-Letter Symbol, and One-Letter Symbol	Structural Formula[a]	Residue Mass (D)[b]	Average Occurrence in Proteins (%)[c]	pK_1 α-COOH[d]	pK_2 α-NH_3^+[d]	pK_R Side Chain[d]
Amino acids with nonpolar side chains						
Glycine Gly G		57.0	7.1	2.35	9.78	
Alanine Ala A		71.1	8.3	2.35	9.87	
Valine Val V		99.1	6.9	2.29	9.74	
Leucine Leu L		113.2	9.7	2.33	9.74	
Isoleucine Ile I		113.2	6.0	2.32	9.76	
Methionine Met M		131.2	2.4	2.13	9.28	
Proline Pro P		97.1	4.7	1.95	10.64	
Phenylalanine Phe F		147.2	3.9	2.20	9.31	
Tryptophan Trp W		186.2	1.1	2.46	9.41	

(*continued*)

[a]The ionic forms shown are those predominating at pH 7.0 (except for that of histidine[e]), although residue mass is given for the neutral compound. The C_α atoms, as well as those atoms marked with an asterisk, are chiral centers with configurations as indicated according to Fischer projection formulas. The standard organic numbering system is provided for heterocycles.

[b]The residue masses are given for the neutral residues. For molecular masses of the parent amino acids, add 18.0 D, the molecular mass of H_2O, to the residue masses. For side chain masses, subtract 56.0 D, the formula mass of a peptide group, from the residue masses.

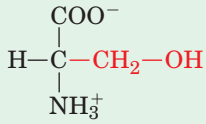
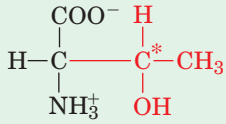
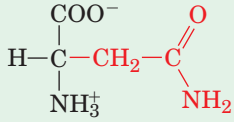
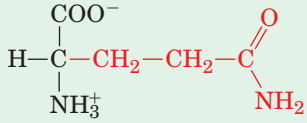
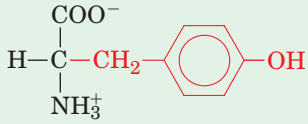
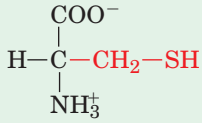
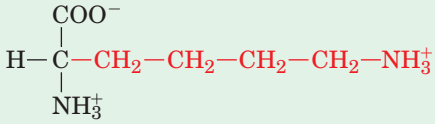
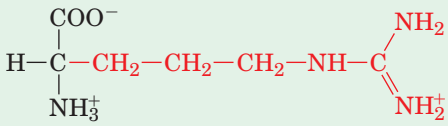
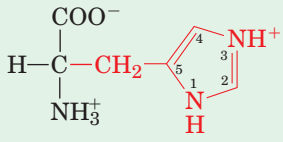
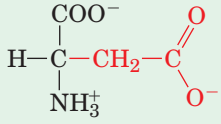
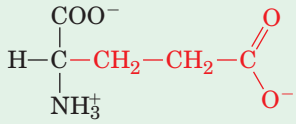
[c]The average amino acid composition in the complete SWISS-PROT database (http://www.expasy.ch/sprot/relnotes/relstat.html), Release 55.11.

[d]From Dawson, R.M.C., Elliott, D.C., Elliott, W.H., and Jones, K.M., *Data for Biochemical Research* (3rd ed.), *pp*. 1–31, Oxford Science Publications (1986).

[e]Both the neutral and protonated forms of histidine are present at pH 7.0 because its pK_R is close to 7.0. The imidazole ring of histidine is numbered here according to the biochemistry convention. In the IUPAC convention, N3 of the biochemistry convention is designated N1 and the numbering increases clockwise around the ring.

[f]The three- and one-letter symbols for asparagine *or* aspartic acid are Asx and B, whereas for glutamine *or* glutamic acid they are Glx and Z. The one-letter symbol for an undetermined *or* "nonstandard" amino acid is X.

Table 4-1 **(*Continued*)**

Name Three-Letter Symbol, and One-Letter Symbol	Structural Formula[a]	Residue Mass (D)[b]	Average Occurrence in Proteins (%)[c]	pK_1 α-COOH[d]	pK_2 α-$^{d}NH_3^{+}$[d]	pK_R Side Chain[d]
Amino acids with uncharged polar side chains						
Serine Ser S		87.1	6.5	2.19	9.21	
Threonine Thr T		101.1	5.3	2.09	9.10	
Asparagine[f] Asn N		114.1	4.0	2.14	8.72	
Glutamine[f] Gln Q		128.1	3.9	2.17	9.13	
Tyrosine Tyr Y		163.2	2.9	2.20	9.21	10.46 (phenol)
Cysteine Cys C		103.1	1.4	1.92	10.70	8.37 (sulfhydryl)
Amino acids with charged polar side chains						
Lysine Lys K		128.2	5.9	2.16	9.06	10.54 (ε-NH_3^{+})
Arginine Arg R		156.2	5.5	1.82	8.99	12.48 (guanidino)
Histidine[e] His H		137.1	2.3	1.80	9.33	6.04 (imidazole)
Aspartic acid[f] Asp D		115.1	5.4	1.99	9.90	3.90 (β-COOH)
Glutamic acid[f] Glu E		129.1	6.8	2.10	9.47	4.07 (γ-COOH)

Molecules that bear charged groups of opposite polarity are known as **zwitterions** (German: *zwitter*, hybrid) or **dipolar ions.** The zwitterionic character of the α-amino acids has been established by several methods including spectroscopic measurements and X-ray crystal structure determinations (in the solid state the α-amino acids are zwitterionic because the basic amine group abstracts a proton from the nearby acidic carboxylic acid group). Because amino acids are zwitterions, their physical properties are characteristic of ionic compounds. For instance, most α-amino acids have melting points near 300°C, whereas their nonionic derivatives usually melt around 100°C. Furthermore, amino acids, like other ionic compounds, are more soluble in polar solvents than in nonpolar solvents. Indeed, most α-amino acids are very soluble in water but are largely insoluble in most organic solvents.

B. *Peptide Bonds*

The α-amino acids polymerize, at least conceptually, through the elimination of a water molecule as is indicated in Fig. 4-3. The resulting CO—NH linkage, which was independently characterized in 1902 by Emil Fischer and Franz Hofmeister, is known as a **peptide bond.** Polymers composed of two, three, a few (3–10), and many **amino acid residues** (alternatively called **peptide units**) are known, respectively, as **dipeptides, tripeptides, oligopeptides,** and **polypeptides.** These substances, however, are often referred to simply as "peptides." *Proteins are molecules that consist of one or more polypeptide chains*. These polypeptides range in length from ~40 to ~34,000 amino acid residues (although few have more than 1500 residues) and, since the average mass of an amino acid residue is ~110 D, have molecular masses that range from ~40 to over ~3700 kD.

Polypeptides are ***linear polymers;*** that is, each amino acid residue is linked to its neighbors in a head-to-tail fashion rather than forming branched chains. This observation reflects the underlying elegant simplicity of the way living systems construct these macromolecules for, as we shall see, the nucleic acids that encode the amino acid sequences of polypeptides are also linear polymers. This permits the direct correspondence between the monomer (nucleotide) sequence of a nucleic acid and the monomer (amino acid) sequence of the corresponding polypeptide without the added complication of specifying the positions and sequences of any branching chains.

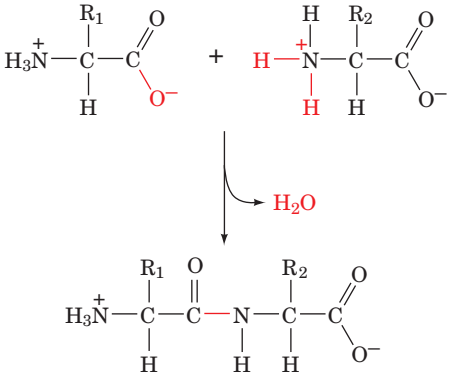


Figure 4-3 **Condensation of two α-amino acids to form a dipeptide.** The peptide bond is shown in red.

With 20 different choices available for each amino acid residue in a polypeptide chain, it is easy to see that a huge number of different protein molecules can exist. For example, for dipeptides, each of the 20 different choices for the first amino acid residue can have 20 different choices for the second amino acid residue, for a total of $20^2 = 400$ distinct dipeptides. Similarly, for tripeptides, there are 20 possibilities for each of the 400 choices of dipeptides to yield a total of $20^3 = 8000$ different tripeptides. A relatively small protein molecule consists of a single polypeptide chain of 100 residues. There are $20^{100} = 1.27 \times 10^{130}$ possible unique polypeptide chains of this length, a quantity vastly greater than the estimated number of atoms in the universe (9×10^{78}). Clearly, nature can have made only a tiny fraction of the possible different protein molecules. Nevertheless, *the various organisms on Earth collectively synthesize an enormous number of different protein molecules whose great range of physicochemical characteristics stem largely from the varied properties of the 20 "standard" amino acids*.

C. *Classification and Characteristics*

The most common and perhaps the most useful way of classifying the 20 "standard" amino acids is according to the polarities of their side chains **(R groups).** This is because proteins fold to their native conformations largely in response to the tendency to remove their hydrophobic side chains from contact with water and to solvate their hydrophilic side chains (Chapters 8 and 9). According to this classification scheme, there are three major types of amino acids: (1) those with nonpolar R groups, (2) those with uncharged polar R groups, and (3) those with charged polar R groups.

a. The Nonpolar Amino Acid Side Chains Have a Variety of Shapes and Sizes

Nine amino acids are classified as having nonpolar side chains. **Glycine** (which, when it was found to be a component of gelatin in 1820, was the first amino acid to be identified in protein hydrolyzates) has the smallest possible side chain, an H atom. **Alanine** (Fig. 4-4), **valine, leucine,** and **isoleucine** have aliphatic hydrocarbon side chains ranging in size from a methyl group for alanine to isomeric butyl groups for leucine and isoleucine. **Methionine** has a thiol ether side chain that resembles an *n*-butyl group in many of its physical properties (C and S have nearly equal electronegativities and S is about the size of a methylene group). **Proline,** a cyclic secondary amino acid, has conformational constraints imposed by the cyclic nature of its pyrrolidine side chain, which is unique among the "standard" 20 amino acids. **Phenylalanine,** with its phenyl moiety (Fig. 4-4), and **tryptophan,** with its indole group, contain

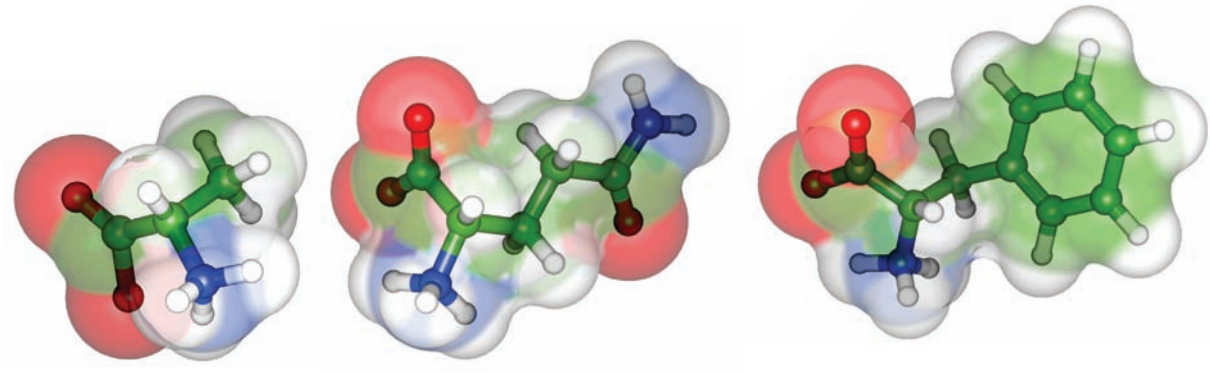


Figure 4-4 **Structures of the α-amino acids alanine, glutamine, and phenylalanine.** The amino acids are shown as ball-and-stick models embedded in their transparent space-filling models. The atoms are colored according to type with C green, H white, N blue, and O red.

aromatic side chains, which are characterized by bulk as well as nonpolarity.

b. Uncharged Polar Side Chains Have Hydroxyl, Amide, or Thiol Groups

Six amino acids are commonly classified as having uncharged polar side chains. **Serine** and **threonine** bear hydroxylic R groups of different sizes. **Asparagine** and **glutamine** (Fig. 4-4) have amide-bearing side chains of different sizes. **Tyrosine** has a phenolic group, which, together with the aromatic groups of phenylalanine and tryptophan, accounts for most of the UV absorbance and fluorescence exhibited by proteins (Section 9-1Cb). **Cysteine** has a thiol group that is unique among the 20 amino acids in that it often forms a disulfide bond to another cysteine residue through the oxidation of their thiol groups (Fig. 4-5). This disulfide bond has great importance in protein structure: *It can join separate polypeptide chains or cross-link two cysteines in the same chain*. Two disulfide-linked cysteines are referred to in the older biochemical literature as the amino acid **cystine** because they were originally thought to form a unique amino acid. However, the discovery that cystine residues arise through the cross-linking of two cysteine residues after polypeptide biosynthesis has occurred has caused the name cystine to become less commonly used.

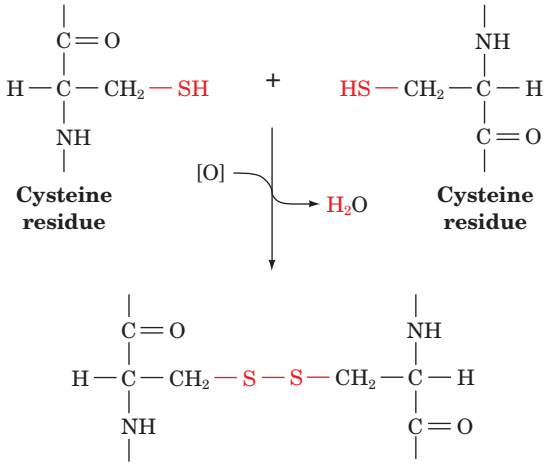


Figure 4-5 **The reaction linking two cysteine residues by a disulfide bond.**

c. Charged Polar Side Chains May Be Positively or Negatively Charged

Five amino acids have charged side chains. The basic amino acids are positively charged at physiological pH values; they are **lysine,** which has a butylammonium side chain, **arginine,** which bears a guanidino group, and **histidine,** which carries an imidazolium moiety. Of the 20 α-amino acids, only histidine, with $pK_R = 6.0$, ionizes within the physiological pH range. At pH 6.0, its imidazole side group is only 50% charged so that histidine is neutral at the basic end of the physiological pH range. As a consequence, histidine side chains often participate in the catalytic reactions of enzymes. The acidic amino acids, **aspartic acid** and **glutamic acid,** are negatively charged above pH 3; in their ionized state, they are often referred to as **aspartate** and **glutamate.** Asparagine and glutamine are, respectively, the amides of aspartic acid and glutamic acid.

The allocation of the 20 amino acids among the three different groups is, of course, somewhat arbitrary. For example, glycine and alanine, the smallest of the amino acids, and tryptophan, with its heterocyclic ring, might just as well be classified as uncharged polar amino acids. Similarly, tyrosine and cysteine, with their ionizable side chains, might also be thought of as charged polar amino acids, particularly at higher pH's, whereas asparagine and glutamine are nearly as polar as their corresponding carboxylates, aspartate and glutamate.

The 20 amino acids vary considerably in their physicochemical properties such as polarity, acidity, basicity, aromaticity, bulk, conformational flexibility, ability to cross-link, ability to hydrogen bond, and chemical reactivity. These several characteristics, many of which are interrelated, are largely responsible for proteins' great range of properties.

D. *Acid–Base Properties*

Amino acids and proteins have conspicuous acid–base properties. The α-amino acids have two or, for those with ionizable side groups, three acid–base groups. The titration curve of glycine, the simplest amino acid, is shown in Fig. 4-6. At low pH values, both acid–base groups of glycine are fully protonated so that it assumes the cationic form $^{+}H_3NCH_2COOH$. In the course of the titration with a strong base, such as NaOH, glycine loses two protons in the stepwise fashion characteristic of a polyprotic acid.

The pK values of glycine's two ionizable groups are sufficiently different so that the Henderson–Hasselbalch equation:

$$\text{pH} = \text{p}K + \log\left(\frac{[\text{A}^-]}{[\text{HA}]}\right) \qquad [2.6]$$

closely approximates each leg of its titration curve. Consequently, the pK for each ionization step is that of the midpoint of its corresponding leg of the titration curve (Sections 2-2A & 2-2C): At pH 2.35 the concentrations of the cationic form, $^{+}H_3NCH_2COOH$, and the zwitterionic form, $^{+}H_3NCH_2COO^{-}$, are equal; similarly, at pH 9.78 the concentrations of the zwitterionic form and the anionic form, $H_2NCH_2COO^{-}$, are equal. Note that *amino acids never assume the neutral form in aqueous solution*.

The pH at which a molecule carries no net electric charge is known as its **isoelectric point, p*I*.** For the α-amino acids, the application of the Henderson–Hasselbalch equation indicates that, to a high degree of precision,

$$\text{p}I = \frac{1}{2}(\text{p}K_i + \text{p}K_j) \qquad [4.1]$$

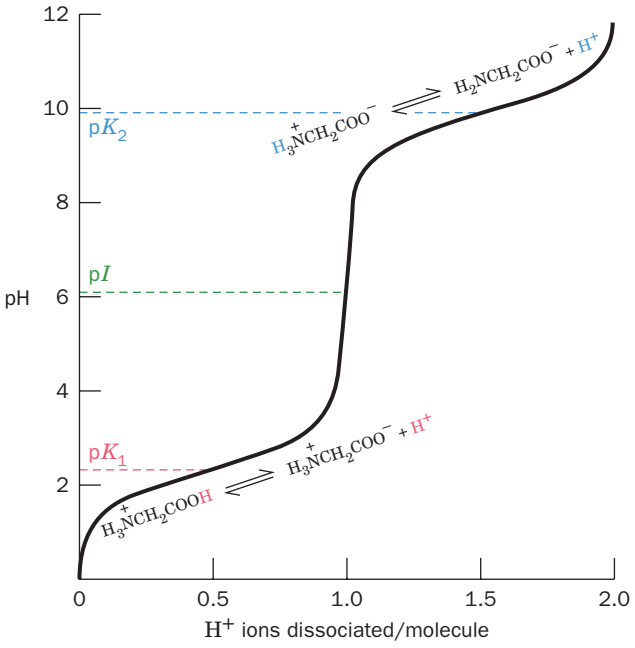


Figure 4-6 **Titration curve of glycine.** Other monoamino, monocarboxylic acids ionize in a similar fashion. [After Meister, A., *Biochemistry of the Amino Acids* (2nd ed.), Vol. 1, *p*. 30, Academic Press (1965).] **See the Animated Figures**

where K_i and K_j are the dissociation constants of the two ionizations involving the neutral species. For monoamino, monocarboxylic acids such as glycine, K_i and K_j represent K_1 and K_2. However, for aspartic and glutamic acids, K_i and K_j are K_1 and K_R, whereas for arginine, histidine, and lysine, these quantities are K_R and K_2.

Acetic acid's pK (4.76), which is typical of aliphatic monocarboxylic acids, is ~2.4 pH units higher than the pK_1 of its α-amino derivative glycine. This large difference in pK values of the same functional group is caused, as is discussed in Section 2-2C, by the electrostatic influence of glycine's positively charged ammonium group; that is, its $—NH_3^+$ group helps repel the proton from its COOH group. Conversely, glycine's carboxylate group increases the basicity of its amino group (pK_2 = 9.78) with respect to that of glycine methyl ester (pK = 7.75). However, the $—NH_3^+$ groups of glycine and its esters are significantly more acidic than are aliphatic amines (p$K \approx 10.7$) because of the electron-withdrawing character of the carboxyl group.

The electronic influence of one functional group on another is rapidly attenuated as the distance between the groups increases. Hence, the pK values of the α-carboxylate groups of amino acids and the side chain carboxylates of aspartic and glutamic acids form a series that is progressively closer in value to the pK of an aliphatic monocarboxylic acid. Likewise, the ionization constant of lysine's side chain amino group is indistinguishable from that of an aliphatic amine.

a. Proteins Have Complex Titration Curves

The titration curves of the α-amino acids with ionizable side chains, such as that of glutamic acid, exhibit the expected three pK values. However, the titration curves of polypeptides and proteins, an example of which is shown in Fig. 4-7, rarely provide any indication of individual pK values because of the large numbers of ionizable groups they represent (typically 30% of a protein's amino acid side chains are ionizable; Table 4-1). Furthermore, the covalent and three-dimensional structure of a protein may cause the pK of each ionizable group to shift by as much as several pH units from its value in the free α-amino acid as a result of the electrostatic influence of nearby charged groups, medium effects arising from the proximity of groups of low dielectric constant, and the effects of hydrogen bonding associations. The titration curve of a protein is also a function of the salt concentration, as is shown in Fig. 4-7, because the salt ions act electrostatically to shield the side chain charges from one another, thereby attenuating these charge–charge interactions.

E. *A Few Words on Nomenclature*

The three-letter abbreviations for the 20 amino acid residues are given in Table 4-1. It is worthwhile memorizing these symbols because they are widely used throughout the biochemical literature, including this text. These abbreviations are, in most cases, taken from the first three letters of the corresponding amino acid's name; they are conversationally pronounced as read.

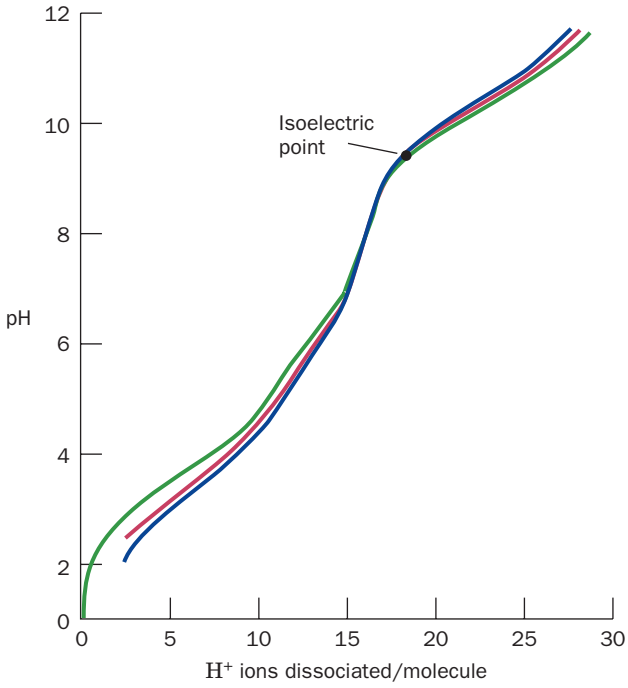


Figure 4-7 **Titration curves of the enzyme ribonuclease A at 25°C.** The concentration of KCl is 0.01*M* for the blue curve, 0.03*M* for the red curve, and 0.15*M* for the green curve. [After Tanford, C. and Hauenstein, J.D., *J. Am. Chem. Soc.* **78,** 5287 (1956).]

The symbol **Glx** means Glu or Gln and, similarly, **Asx** means Asp or Asn. These ambiguous symbols stem from laboratory experience: Asn and Gln are easily hydrolyzed to aspartic acid and glutamic acid, respectively, under the acidic or basic conditions that are usually used to excise them from proteins. Therefore, without special precautions, we cannot determine whether a detected Glu was originally Glu or Gln, and likewise for Asp and Asn.

The one-letter symbols for the amino acids are also given in Table 4-1. This more compact code is often used when comparing the amino acid sequences of several similar proteins and hence should also be memorized. Note that the one-letter symbols are usually the first letter of the amino acid residue's name. However, for those sets of residues that have the same first letter, this is only true of the most abundant residue of the set.

Amino acid residues in polypeptides are named by dropping the suffix **-ine** in the name of the amino acid and replacing it by **-yl.** Polypeptide chains are described by starting at the amino terminus (known as the **N-terminus**) and sequentially naming each residue until the carboxyl terminus (the **C-terminus**) is reached. The amino acid at the C-terminus is given the name of its parent amino acid. Thus the compound shown in Fig. 4-8 is alanyltyrosylaspartylglycine. Of course such names for polypeptide chains of more than a few residues are extremely cumbersome. The use of abbreviations for amino acid residues partially relieves this problem. Thus the foregoing tetrapeptide is Ala-Tyr-Asp-Gly using the three-letter abbreviations and AYDG using the one-letter symbols. Note that these abbreviations are always written so that the N-terminus of the polypeptide chain is to the left and the C-terminus is to the right.

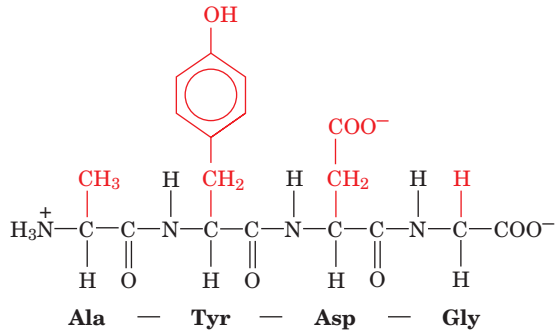


Figure 4-8 **The tetrapeptide Ala-Tyr-Asp-Gly.**

The various nonhydrogen atoms of the amino acid side chains are often named in sequence with the Greek alphabet (α, β, γ, δ, ε, ζ, η, …) starting at the carbon atom adjacent to the peptide carbonyl group (the C_α atom). Therefore, as Fig. 4-9 indicates, Glu has a γ-carboxyl group and Lys has a ζ-amino group (alternatively known as an ε-amino group because the N atom is substituent to C_ε). Unfortunately, this labeling system is ambiguous for several amino acids. Consequently, standard numbering schemes for organic molecules are also employed. These are indicated in Table 4-1 for the heterocyclic side chains.

2 OPTICAL ACTIVITY

The amino acids as isolated by the mild hydrolysis of proteins are, with the exception of glycine, all **optically active;** that is, they rotate the plane of plane-polarized light (see below).

Optically active molecules have an asymmetry such that they are not superimposable on their mirror image in the same way that a left hand is not superimposable on its mirror image, a right hand. This situation is characteristic of substances that contain tetrahedral carbon atoms that have four different substituents. The two such molecules

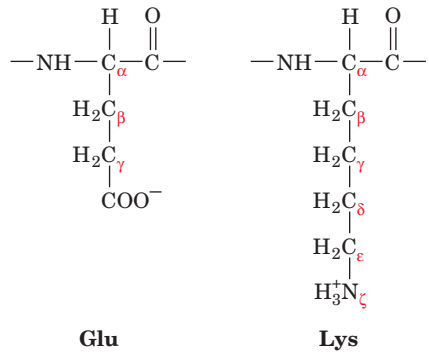


Figure 4-9 **Greek lettering scheme used to identify the atoms in the glutamyl and lysyl R groups.**

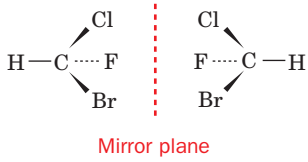


Figure 4-10 **The two enantiomers of fluorochlorobromomethane.** The four substituents are tetrahedrally arranged about the central atom with the dotted lines indicating that a substituent lies behind the plane of the paper, a triangular line indicating that it lies above the plane of the paper, and a thin line indicating that it lies in the plane of the paper. The mirror plane relating the enantiomers is represented by a vertical dashed line.

depicted in Fig. 4-10 are not superimposable since they are mirror images. The central atoms in such atomic constellations are known as **asymmetric centers** or **chiral centers** and are said to have the property of **chirality** (Greek: *cheir,* hand). The C_α atoms of all the amino acids, with the exception of glycine, are asymmetric centers. Glycine, which has two H atoms substituent to its C_α atom, is superimposable on its mirror image and is therefore not optically active.

Molecules that are nonsuperimposable mirror images are known as **enantiomers** of one another. Enantiomeric molecules are physically and chemically indistinguishable by most techniques. *Only when probed asymmetrically, for example, by plane-polarized light or by reactants that also contain chiral centers, can they be distinguished and/or differentially manipulated.*

There are three commonly used systems of nomenclature whereby a particular stereoisomer of an optically active molecule can be classified. These are explained in the following sections.

A. *An Operational Classification*

Molecules are classified as **dextrorotatory** (Greek: *dexter,* right) or **levorotatory** (Greek: *laevus,* left) depending on whether they rotate the plane of plane-polarized light clockwise or counterclockwise from the point of view of the observer. This can be determined by an instrument known as a **polarimeter** (Fig. 4-11). A quantitative measure of the optical activity of the molecule is known as its **specific rotation:**

$$[\alpha]_D^{25} = \frac{\text{observed rotation (degrees)}}{\text{optical path length (dm)} \times \text{concentration } (g \cdot cm^{-3})} \qquad [4.2]$$

where the superscript 25 refers to the temperature at which polarimeter measurements are customarily made (25°C) and the subscript D indicates the monochromatic light that is traditionally employed in polarimetry, the so-called D-line in the spectrum of sodium (589.3 nm). Dextrorotatory and levorotatory molecules are assigned positive and negative values of $[\alpha]_D^{25}$. Dextrorotatory molecules are therefore designated by the prefix (+) and their levorotatory enantiomers have the prefix (−). In an equivalent but archaic nomenclature, the lowercase letters *d (dextro)* and *l (levo)* are used.

The sign and magnitude of a molecule's specific rotation depend on the structure of the molecule in a complicated and poorly understood manner. It is not yet possible to predict reliably the magnitude or even the sign of a given molecule's specific rotation. For example, proline, leucine, and arginine, which are isolated from proteins, have specific rotations in pure aqueous solutions of −86.2°, −10.4°, and +12.5°, respectively. Their enantiomers exhibit values of $[\alpha]_D^{25}$ of the same magnitude but of opposite signs. As might be expected from the acid–base nature of the amino acids, these quantities vary with the solution pH.

A problem with this operational classification system for optical isomers is that it provides no presently interpretable indication of the **absolute configuration** (spatial arrangement) of the chemical groups about a chiral center. Furthermore, a molecule with more than one asymmetric center may have an optical rotation that is not obviously

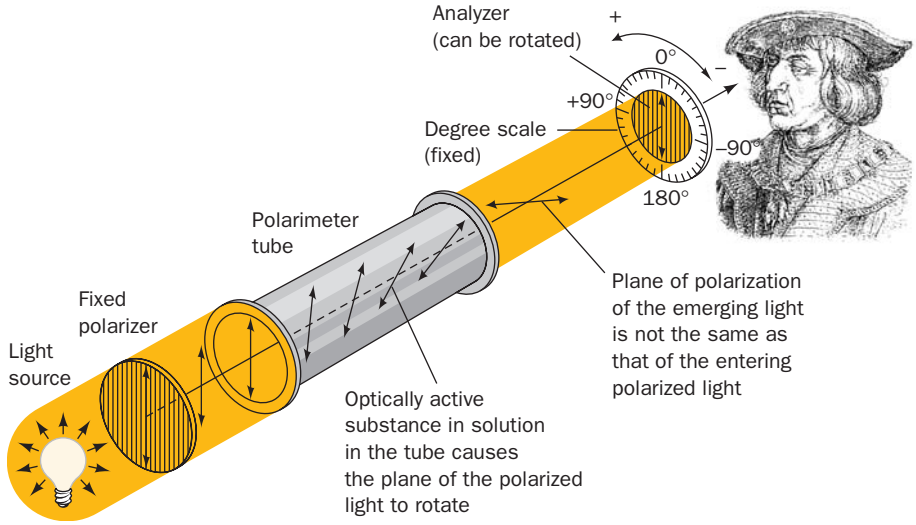


Figure 4-11 **Schematic diagram of a polarimeter.** This device is used to measure optical rotation.

related to the rotatory powers of the individual asymmetric centers. For this reason, the following relative classification scheme is more useful.

B. *The Fischer Convention*

In this system, the configuration of the groups about an asymmetric center is related to that of **glyceraldehyde,** a molecule with one asymmetric center. By a convention introduced by Fischer in 1891, the (+) and (−) stereoisomers of glyceraldehyde are designated **D-glyceraldehyde** and **L-glyceraldehyde,** respectively (note the use of small uppercase letters). With the realization that there was only a 50% chance that he was correct, Fischer assumed that the configurations of these molecules were those shown in Fig. 4-12. Fischer also proposed a convenient shorthand notation for these molecules, known as **Fischer projections,** which are also given in Fig. 4-12. In the Fischer convention, horizontal bonds extend above the plane of the paper and vertical bonds extend below the plane of the paper as is explicitly indicated by the accompanying geometrical formulas.

The configuration of groups about a chiral center can be related to that of glyceraldehyde by chemically converting these groups to those of glyceraldehyde using reactions of known stereochemistry. For α-amino acids, the arrangement of the amino, carboxyl, R, and H groups about the C_α atom is related to that of the hydroxyl, aldehyde, CH_2OH, and H groups, respectively, of glyceraldehyde. In this way, L-glyceraldehyde and L-α-amino acids are said to have the same relative configurations (Fig. 4-13). Through the use of this method, the configurations of the

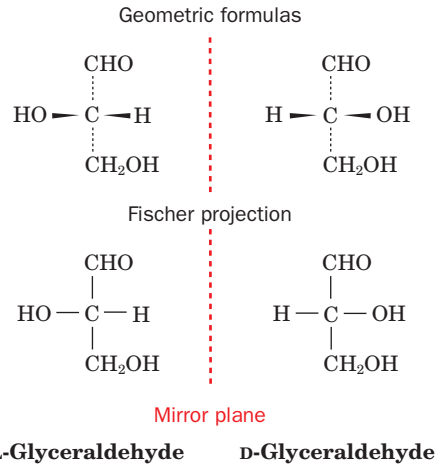


Figure 4-12 **Fischer convention configurations for naming the enantiomers of glyceraldehyde.** Glyceraldehyde enantiomers are represented by geometric formulas (*top*) and their corresponding Fischer projection formulas (*bottom*). Note that in Fischer projections, all horizontal bonds point above the page and all vertical bonds point below the page. The mirror planes relating the enantiomers are represented by a vertical dashed line. (Fischer projection formulas, as traditionally presented, omit the central C symbolizing the chiral carbon atom. The Fischer projection formulas in this text, however, will generally have a central C.)

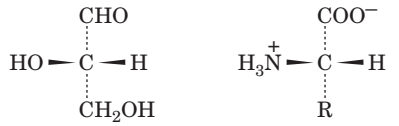


Figure 4-13 **Configurations of L-glyceraldehyde and L-α-amino acids.**

α-amino acids can be described without reference to their specific rotations.

All α-amino acids derived from proteins have the L stereochemical configuration; that is, they all have the same relative configuration about their C_α atoms. In 1949, it was demonstrated by a then new technique in X-ray crystallography that Fischer's arbitrary choice was correct: The designation of the relative configuration of chiral centers is the same as their absolute configuration. The absolute configuration of L-α-amino acid residues may be easily remembered through the use of the "CORN crib" mnemonic that is diagrammed in Fig. 4-14.

a. Diastereomers Are Chemically and Physically Distinguishable

A molecule may have multiple asymmetric centers. For such molecules, the terms **stereoisomers** and **optical isomers** refer to molecules with different configurations about at least one of their chiral centers, but that are otherwise identical. The term enantiomer still refers to a molecule that is the mirror image of the one under consideration, that is, different in all its chiral centers. Since each asymmetric center in a chiral molecule can have two possible configurations, a molecule with n chiral centers has 2^n different possible stereoisomers and 2^{n-1} enantiomeric pairs. Threonine and isoleucine each have two chiral centers and hence $2^2 = 4$ possible stereoisomers. The forms of threonine and isoleucine that are isolated from proteins, which are by convention called the L forms, are indicated in Table 4-1. The mirror images of the L forms are the D forms. Their other two optical isomers are said to be **diastereomers** (or **allo** forms) of the enantiomeric D and L forms. The relative

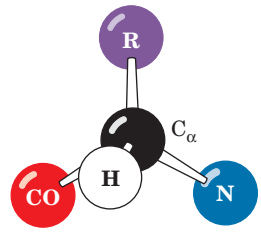


Figure 4-14 **"CORN crib" mnemonic for the hand of L-amino acids.** Looking at the C_α atom from its H atom substituent, its other substituents should read CO—R—N in the clockwise direction as shown. Here CO, R, and N, respectively, represent the carbonyl group, side chain, and main chain nitrogen atom. [After Richardson, J.S., *Adv. Protein Chem.* **34,** 171 (1981).]

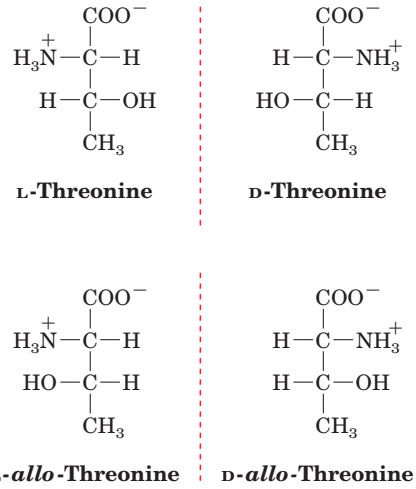


Figure 4-15 **Fischer projections of threonine's four stereoisomers.** The D and L forms are mirror images as are the D-*allo* and L-*allo* forms. D- and L-threonine are each diastereomers of both D-*allo*- and L-*allo*-threonine.

configurations of all four stereoisomers of threonine are given in Fig. 4-15. Note the following points:

1. The D-*allo* and L-*allo* forms are mirror images of each other, as are the D and L forms. Neither allo form is symmetrically related to either of the D or L forms.

2. In contrast to the case for enantiomeric pairs, diastereomers are physically and chemically distinguishable from one another by ordinary means such as melting points, spectra, and chemical reactivity; that is, they are really different compounds in the usual sense.

A special case of diastereoisomerism occurs when the two asymmetric centers are chemically identical. Two of the four Fischer projections of the sort shown in Fig. 4-15 then represent the same molecule. This is because the two asymmetric centers in this molecule are mirror images of each other. Such a molecule is superimposible on its mirror image and is therefore optically inactive. This so-called **meso** form is said to be **internally compensated.** The three optical isomers of cystine are shown in Fig. 4-16, where it can be seen that the D and L isomers are mirror images of each other as before. Only L-cystine occurs in proteins.

C. *The Cahn–Ingold–Prelog System*

Despite its usefulness, the Fischer scheme is awkward and often ambiguous for molecules with more than one asymmetric center. For this reason, the following absolute nomenclature scheme was formulated in 1956 by Robert Cahn, Christopher Ingold, and Vladimir Prelog. In this system, the four groups surrounding a chiral center are ranked according to a specific although arbitrary priority scheme: *Atoms of higher atomic number bonded to a chiral center are ranked above those of lower atomic number*. For example, the oxygen atom of an OH group takes precedence over the carbon atom of a CH_3 group that is bonded to the same chiral C atom. If any of the first substituent atoms are of the same element, the priority of these groups is established from the atomic numbers of the second, third, etc., atoms outward from the asymmetric center. Hence a CH_2OH group takes precedence over a CH_3 group. There are other rules (given in the references and in many organic chemistry textbooks) for assigning priority ratings to substituents with multiple bonds or differing isotopes. The order of priority of some common functional groups is

$$SH > OH > NH_2 > COOH > CHO > CH_2OH > C_6H_5 > CH_3 > {}^2H > {}^1H$$

Note that each of the groups substituent to a chiral center must have a different priority rating; otherwise the center could not be asymmetric.

The prioritized groups are assigned the letters W, X, Y, Z such that their order of priority rating is W > X > Y > Z. To establish the configuration of the chiral center, it is viewed from the asymmetric center toward the Z group (lowest priority). *If the order of the groups* W → X → Y *as seen from this direction is clockwise, then the configuration of the asymmetric center is designated (R)* (Latin: *rectus,* right). *If the order of* W → X → Y *is counterclockwise, the asymmetric center is designated (S)* (Latin: *sinister,* left). L-Glyceraldehyde is therefore designated (*S*)-glyceraldehyde (Fig. 4-17) and, similarly, L-alanine is (*S*)-alanine (Fig. 4-18). In fact, all the L-amino acids from proteins are (*S*)-amino acids, with the exception of L-cysteine, which is (*R*)-cysteine.

A major advantage of this so-called **Cahn–Ingold– Prelog** or **(*RS*) system** is that the chiralities of compounds with multiple asymmetric centers can be unambiguously described. Thus, in the (*RS*) system, L-threonine is (2*S*,3*R*)-threonine, whereas L-isoleucine is (2*S*,3*S*)-isoleucine (Fig. 4-19).

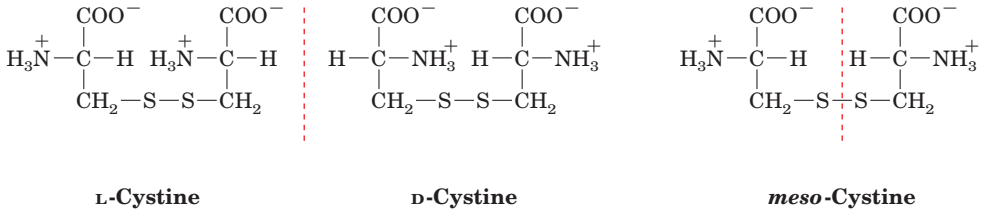
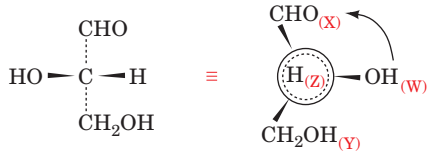
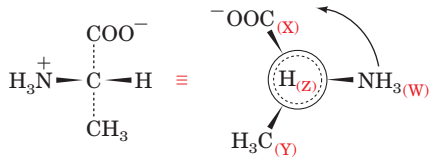


Figure 4-16 **The three stereoisomers of cystine.** The D and L forms are related by mirror symmetry, whereas the meso form has internal mirror symmetry and therefore lacks optical activity.



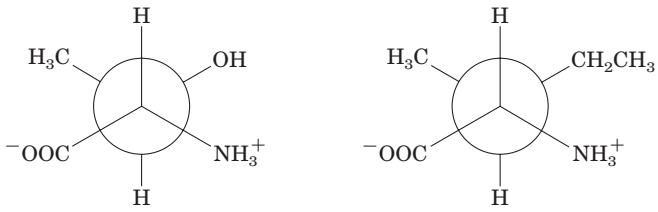
L-Glyceraldehyde **(*S*)-Glyceraldehyde**

Figure 4-17 **The structural formula of L-glyceraldehyde.** Its equivalent (*RS*) system representation indicates that it is (*S*)-glyceraldehyde. In the latter drawing, the chiral C atom is represented by the large circle, and the H atom, which is located behind the plane of the paper, is represented by the smaller concentric dashed circle.



L-Alanine **(*S*)-Alanine**

Figure 4-18 **The structural formula of L-alanine.** Its equivalent (*RS*) system representation indicates that it is (*S*)-alanine.



(2*S*,3*R*)-Threonine **(2*S*,3*S*)-Isoleucine**

Figure 4-19 **Newman projection diagrams of the stereoisomers of threonine and isoleucine derived from proteins.** Here the C_α—C_β bond is viewed end on. The nearer atom, C_α, is represented by the confluence of the three bonds to its substituents, whereas the more distant atom, C_β, is represented by a circle from which its three substituents project.

a. Prochiral Centers Have Distinguishable Substituents

Two chemically identical substituents to an otherwise chiral tetrahedral center are geometrically distinct; that is, the center has no rotational symmetry so that it can be unambiguously assigned left and right sides. Consider, for example, the substituents to the C1 atom of ethanol (the CH_2 group; Fig. 4-20*a*). If one of the H atoms were converted to another group (not CH_3 or OH), C1 would be a chiral center. The two H atoms are therefore said to be **prochiral.** If we arbitrarily assign the H atoms the subscripts *a* and *b* (Fig. 4-20), then H_b is said to be ***pro-R*** because in sighting from C1 toward H_a (as if it were the Z group of a chiral center), the order of priority of the other substituents decreases in a clockwise direction (Fig. 4-20*b*). Similarly, H_a is said to be ***pro-S*** (Fig. 4-20*c*).

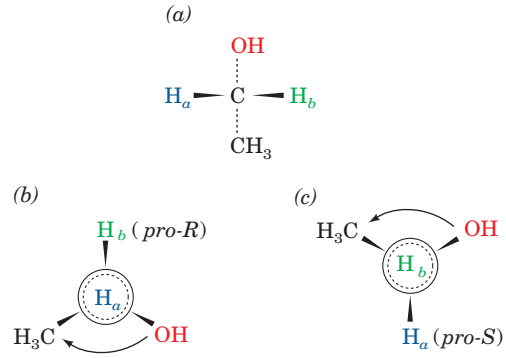


Figure 4-20 **Views of ethanol.** (*a*) Note that H_a and H_b, although chemically identical, are distinguishable: Rotating the molecule by 180° about the vertical axis so as to interchange these two hydrogen atoms does not yield an indistinguishable view of the molecule because the rotation also interchanges the chemically different OH and CH_3 groups. (*b*) Looking from C1 to H_a, the *pro-S* hydrogen atom (the dotted circle). (*c*) Looking from C1 to H_b, the *pro-R* hydrogen atom.

Planar objects with no rotational symmetry also have the property of prochirality. For example, in many enzymatic reactions, stereospecific addition to a trigonal carbon atom occurs from a particular side of that carbon atom to yield a chiral center (Section 13-2A). If a trigonal carbon is facing the viewer such that the order of priority of its substituents decreases in a clockwise manner (Fig. 4-21*a*), that face is designated as the ***re* face** (after *rectus*). The opposite face is designated as the ***si* face** (after *sinister*) since the priorities of its substituents decrease in the counterclockwise direction (Fig. 4-21*b*). Comparison of Figs. 4-20*b* and 4-21*a* indicates that an H atom adding to the *re* side of acetaldehyde atom C1 occupies the *pro-R* position of the resulting tetrahedral center. Conversely, a *pro-S* H atom is generated by *si* side addition to this trigonal center (Figs. 4-20*c* and 4-21*b*).

Closely related compounds that have the same configurational representation under the Fischer DL convention may have different representations under the (*RS*) system. Consequently, we shall use the Fischer convention in most cases. The (*RS*) system, however, is indispensable for describing prochirality and stereospecific reactions, so we shall find it invaluable for describing enzymatic reactions.

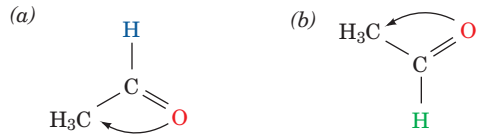


Figure 4-21 **Views of acetaldehyde.** (*a*) Its *re* face and (*b*) its *si* face.

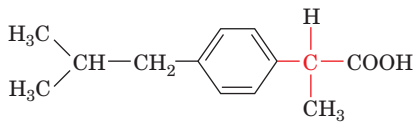
D. *Chirality and Biochemistry*

The ordinary chemical synthesis of chiral molecules produces **racemic** mixtures of these molecules (equal amounts of each member of an enantiomeric pair) because ordinary chemical and physical processes have no stereochemical bias. Consequently, there are equal probabilities for an asymmetric center of either hand to be produced in any such process. In order to obtain a product with net optical activity, a chiral process must be employed. This usually takes the form of using chiral reagents, although, at least in principle, the use of any asymmetric influence such as light that is plane polarized in one direction can produce a net asymmetry in a reaction product.

One of the most striking characteristics of life is its production of optically active molecules. *The biosynthesis of a substance possessing asymmetric centers almost invariably produces a pure stereoisomer.* The fact that the amino acid residues of proteins all have the L configuration is just one example of this phenomenon. This observation has prompted the suggestion that a simple diagnostic test for the past or present existence of extraterrestrial life, be it on moon rocks or in meteorites that have fallen to Earth, would be the detection of net optical activity in these materials. Any such finding would suggest that the asymmetric molecules thereby detected had been biosynthetically produced. Thus, even though α-amino acids have been extracted from carbonaceous meteorites, the observation that they come in racemic mixtures suggests that they are of chemical rather than biological origin.

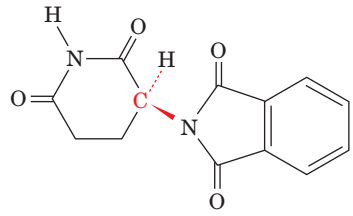
One of the enigmas of the origin of life is why terrestrial life is based on certain chiral molecules rather than their enantiomers, that is, on L-amino acids, for example, rather than D-amino acids. Arguments that physical effects such as polarized light might have promoted significant net asymmetry in prebiotically synthesized molecules (Section 1-5B) have not been convincing. Perhaps L-amino acid–based life-forms arose at random and simply "ate" any D-amino acid–based life-forms.

The importance of stereochemistry in living systems is also a concern of the pharmaceutical industry. *Many drugs are chemically synthesized as racemic mixtures, although only one enantiomer has biological activity.* In most cases, the opposite enantiomer is biologically inert and is therefore packaged along with its active counterpart. This is true, for example, of the widely used anti-inflammatory agent **ibuprofen,** only one enantiomer of which is physiologically active (Fig. 4-22). Occasionally, the inactive enantiomer of a useful drug produces harmful effects and must therefore be eliminated from the racemic mixture. The most striking example of this is the drug **thalidomide** (Fig. 4-23), a mild sedative whose "inactive" enantiomer causes severe birth defects. Partly because of the unanticipated problems caused by "inactive" drug enantiomers, **chiral organic synthesis** has become an active area of medicinal chemistry.



Ibuprofen

Figure 4-22 **Ibuprofen.** Only the enantiomer shown has anti-inflammatory action. The chiral carbon is red.



Thalidomide

Figure 4-23 **Thalidomide.** This drug was widely used in Europe as a mild sedative in the early 1960s. Its inactive enantiomer (not shown), which was present in equal amounts in the formulations used, causes severe birth defects in humans when taken during the first trimester of pregnancy. Thalidomide was often prescribed to alleviate the nausea (morning sickness) that is common during this period.

3 "NONSTANDARD" AMINO ACIDS

The 20 common amino acids are by no means the only amino acids that occur in biological systems. "Nonstandard" amino acid residues are often important constituents of proteins and biologically active polypeptides. Many amino acids, however, are not constituents of proteins. Together with their derivatives, they play a variety of biologically important roles.

A. *Amino Acid Derivatives in Proteins*

The "universal" genetic code, which is nearly identical in all known life-forms (Section 5-4Bb), specifies only the 20 "standard" amino acids of Table 4-1. Nevertheless, many other amino acids, a selection of which is given in Fig. 4-24, are components of certain proteins. *In all known cases but two (Section 32-2De), however, these unusual amino acids result from the specific modification of an amino acid residue after the polypeptide chain has been synthesized.* Among the most prominent of these modified amino acid residues are **4-hydroxyproline** and **5-hydroxylysine.** Both of these amino acid residues are important structural constituents of the fibrous protein **collagen,** the most abundant protein in mammals (Section 8-2B). Amino acids of proteins that form complexes with nucleic acids are often modified. For example, the chromosomal proteins known as **histones** may be specifically methylated, acetylated, and/or phosphorylated at specific Lys, Arg, and Ser residues (Section 34-3Baa). Several of these derivatized

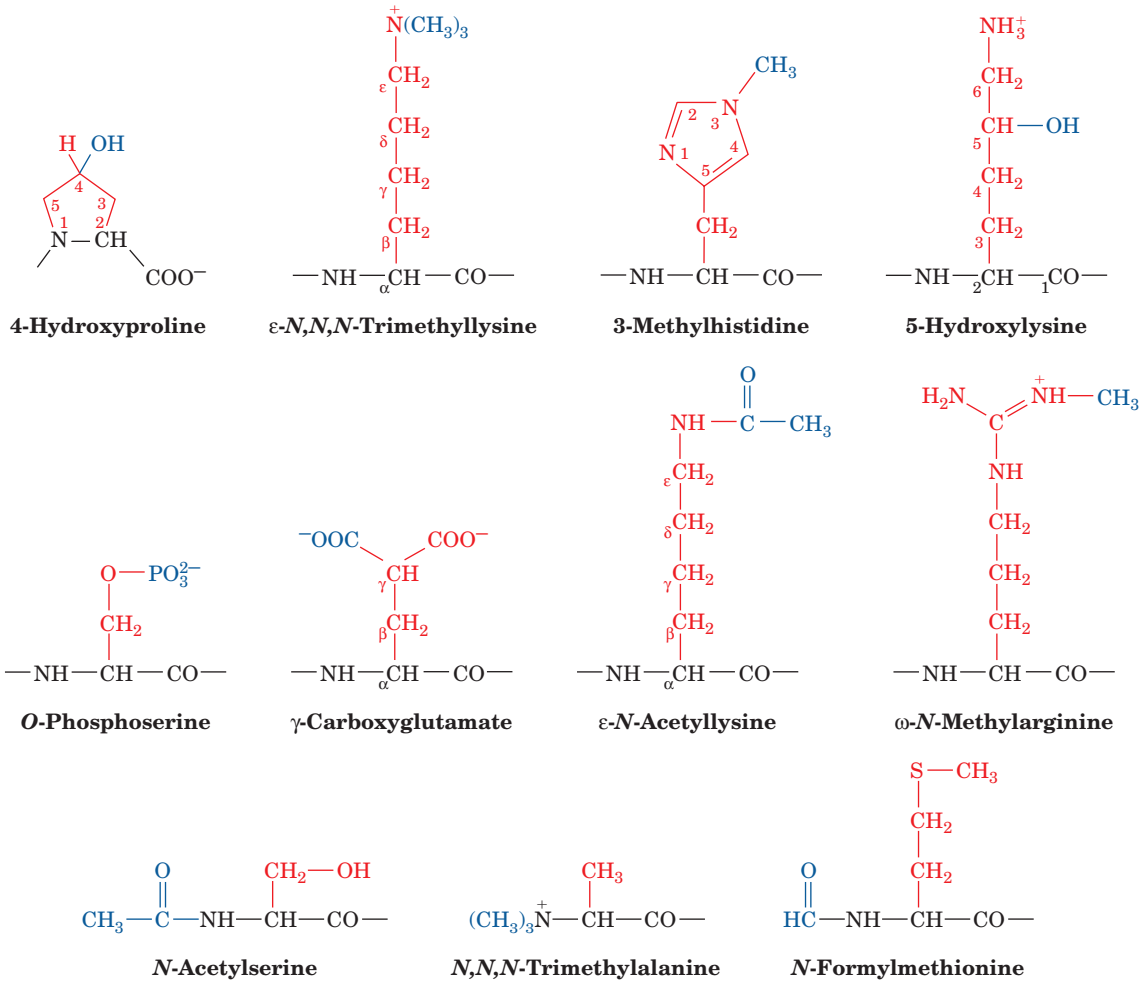


Figure 4-24 **Some uncommon amino acid residues that are components of certain proteins.** All of these residues are modified from one of the 20 "standard" amino acids after polypeptide chain biosynthesis. Those amino acid residues that are derivatized at their N_α position occur at the N-termini of proteins.

amino acid residues are presented in Fig. 4-24. ***N*-Formylmethionine** is initially the N-terminal residue of all prokaryotic proteins, but is usually removed as part of the protein maturation process (Section 32-3Ca). **γ-Carboxyglutamic acid** is a constituent of several proteins involved in blood clotting (Section 35-1Ba). Note that in most cases, these modifications are important, if not essential, for the function of the protein.

D-Amino acid residues are components of many of the relatively short (<20 residues) bacterial polypeptides that are enzymatically rather than ribosomally synthesized. These polypeptides are perhaps most widely distributed as constituents of bacterial cell walls (Section 11-3Ba), which D-amino acids render less susceptible to attack by the **peptidases** (enzymes that hydrolyze peptide bonds) that many organisms employ to digest bacterial cell walls. Likewise, D-amino acids are components of many bacterially produced peptide antibiotics including **valinomycin, gramicidin A** (Section 20-2C), and **actinomycin D** (Section 31-2Cc). D-Amino acid residues are also functionally essential components of several ribosomally synthesized polypeptides of eukaryotic as well as prokaryotic origin. These D-amino acid residues are posttranslationally formed, most probably through the enzymatically mediated inversion of the preexisting L-amino acid residues.

B. *Specialized Roles of Amino Acids*

Besides their role in proteins, amino acids and their derivatives have many biologically important functions. A few examples of these substances are shown in Fig. 4-25. This alternative use of amino acids is an example of the biological opportunism that we shall repeatedly encounter: *Nature tends to adapt materials and processes that are already present to new functions.*

Amino acids and their derivatives often function as chemical messengers in the communications between cells. For example, glycine, **γ-aminobutyric acid (GABA;** a glutamate

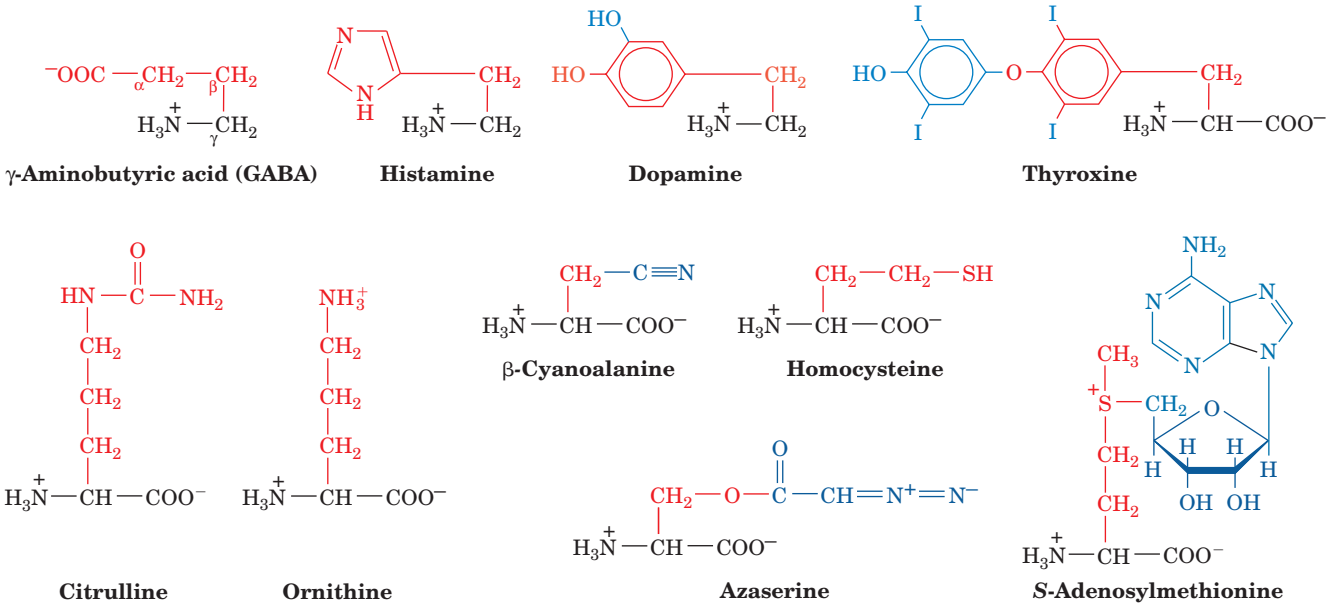


Figure 4-25 **Some biologically produced derivatives of "standard" amino acids and amino acids that are not components of proteins.**

decarboxylation product), and **dopamine** (a tyrosine derivative) are neurotransmitters (substances released by nerve cells to alter the behavior of their neighbors; Section 20-5C); **histamine** (the decarboxylation product of histidine) is a potent local mediator of allergic reactions; and **thyroxine** (a tyrosine derivative) is an iodine-containing thyroid hormone that generally stimulates vertebrate metabolism (Section 19-1D).

Certain amino acids are important intermediates in various metabolic processes. Among them are **citrulline** and **ornithine,** intermediates in urea biosynthesis (Section 26-2B); **homocysteine,** an intermediate in amino acid metabolism (Section 26-3Ea); and ***S*-adenosylmethionine,** a biological methylating reagent (Section 26-3Ea).

Nature's diversity is remarkable. Over 700 different amino acids have been found in various plants, fungi, and bacteria, most of which are α-amino acids. For the most part, their biological roles are obscure although the fact that many are toxic suggests that they have a protective function. Indeed, some of them, such as **azaserine,** are medically useful antibiotics. Many of these amino acids are simple derivatives of the 20 "standard" amino acids although some of them, including azaserine and **β-cyanoalanine** (Fig. 4-25), have unusual structures.

CHAPTER SUMMARY

1 The Amino Acids of Proteins Proteins are linear polymers that are synthesized from the same 20 "standard" α-amino acids through their condensation to form peptide bonds. These amino acids all have a carboxyl group with a pK near 2.2 and an amino substituent with a pK near 9.4 attached to the same carbon atom, the C_α atom. The α-amino acids are zwitterionic compounds, $^+H_3N{-}CHR{-}COO^-$, in the physiological pH range. The various amino acids are usually classified according to the polarities of their side chains, R, which are substituent to the C_α atom. Glycine, alanine, valine, leucine, isoleucine, methionine, proline (which is really a secondary amino acid), phenylalanine, and tryptophan are nonpolar amino acids; serine, threonine, asparagine, glutamine, tyrosine, and cysteine are uncharged and polar; and lysine, arginine, histidine, aspartic acid, and glutamic acid are charged and polar. The side chains of many of these amino acids bear acid–base groups, and hence the properties of the proteins containing them are pH dependent.

2 Optical Activity The C_α atoms of all α-amino acids except glycine each bear four different substituents and are therefore chiral centers. According to the Fischer convention, which relates the configuration of D- or L-glyceraldehyde to that of the asymmetric center of interest, all the amino acids of proteins have the L configuration; that is, they all have the same absolute configuration about their C_α atom. According to the Cahn–Ingold–Prelog (*RS*) system of chirality nomenclature, they are, with the exception of cysteine, all (*S*)-amino acids. The side chains of threonine and isoleucine also contain chiral centers. A prochiral center has no rotational symmetry, and hence its substituents, in the case of a central atom, or its faces, in the case of a planar molecule, are distinguishable.

3 "Nonstandard" Amino Acids Amino acid residues other than the 20 from which proteins are synthesized also have important biological functions. These "nonstandard" residues result from the specific chemical modifications of amino acid residues in preexisting proteins. Amino acids and their derivatives also have independent biological roles such as neurotransmitters, metabolic intermediates, and poisons.